

THE CLAIMS

Please amend the Claims as follows:

Claims 1-24. (Canceled)

25. (Currently amended) A computer implemented method of searching biochemical ~~sequence data or~~ information for a plurality of query sequences in a set of target sequence fragments, allowing for mismatches at up to n sequence positions, comprising:
- 1a) providing a first query distribution scheme by dividing each query sequence into n+1 query sequence segments and
providing a corresponding first target distribution scheme by dividing each target fragment into at least n+1 target sequence fragment segments;
 - 1b) for each of the n+1 query sequences under the first query distribution scheme, constructing a first query group and a second query group according to the first query distribution scheme by placing individual query sequence segments in one of said query groups such so that at least n query sequence segments are contained in the second query group;
 - 1c) for each of the target sequence fragment segments under the first target distribution scheme, constructing a first target group and a second target group according to the first target distribution scheme by placing individual target sequence fragment segments in one of said target groups, same-distribution of segments as that of the first query group, and constructing a second target group according to the same distribution of segments as that of the second query group;
 - 1d) for each of the n+1 query sequences, comparing the first query group with

the corresponding first target group to identify a potential matches under the first query distribution scheme and the corresponding first target distribution scheme;

- ~~e) performing steps b), c), and d), for different distributions of segments; and~~
- 2a) providing a second query distribution scheme by dividing each query sequence into $n+1$ query sequence segments and
providing a corresponding second target distribution scheme by dividing each target fragment into at least $n+1$ target sequence fragment segments;
- 2b) for each of the $n+1$ query sequences under the second query distribution scheme, constructing a first query group and a second query group according to the first query distribution scheme by placing individual query sequence segments in one of said query groups so that at least n query sequence segments are contained in the second query group;
- 2c) for each of the target sequence fragment segments under the second target distribution scheme, constructing a first target group and a second target group according to the first target distribution scheme by placing individual target sequence fragment segments in one of said query groups;
- 2d) for each of the $n+1$ query sequences, comparing the first query group for a sequence identical with the corresponding first target group to identify a potential matches under the second query distribution scheme and the corresponding second target distribution scheme; and
- 3) optionally repeating steps 2a) to 2d) for further query distribution schemes and corresponding target distribution schemes;

~~f) — for each match identified in steps d) and e),~~

wherein for each of the potential matches identified under the first distribution schemes obtained in steps 1d), under the second distribution schemes obtained in step 2d), and in the distribution schemes in step 3), subsequently

4) comparing the second query group with the second target group to identify a match,

thereby identifying a query sequence in the set of target sequence fragments, while allowing for mismatches in up to n sequence positions.

26. (Canceled)

27. (Canceled)

28. (Canceled)

29. (Currently amended) The computer implemented method of claim 25, wherein step (~~f~~ 4) is carried out by applying an exclusive OR operation to a binary representation of each of the second query group and the second target group.

30. (Previously presented) The computer implemented method of claim 29, wherein a result of the exclusive OR operation is analyzed using a lookup table.

31. (Previously presented) The computer implemented method of claim 25, further comprising:
for each distinct distribution of query sequence segments, constructing a first query table indexed by possible values of the first query groups, wherein the entries in the first query table provide access to each second query group by using as an index the value of a corresponding first query

group.

32. (Previously presented) The computer implemented method of claim 31, further comprising:
for each distinct distribution of query sequence segments, constructing a second query table providing access to each second query group, wherein the entries in the first query table provide references to appropriate entries in the second query table.

33. (Currently amended) The computer implemented method of claim 31, further comprising:
for each first target group constructed in step 1(c) or 2c), implementing step 1(d) or 2d) by using each first target group to form an index into the first query table.

34. (Previously presented) The computer implemented method of claim 31, wherein, if two distinct distributions of query sequence segments are such that the first query group of a first distribution is the same as the second query group of a second distribution, respective first query tables for each of the two distinct distributions are at least one of constructed and used concurrently.

35. (Previously presented) The computer implemented method of claim 25, wherein target sequence fragments in the set of target sequence fragments comprise overlapping fragments of one or more target sequences.

36. (Previously presented) The computer implemented method of claim 25, wherein each query sequence of the plurality of query sequences and the target sequence fragments comprise nucleotide sequence data.

37. (Previously presented) The computer implemented method of claim 36, wherein each query sequence of the plurality of query sequences and the target sequence fragments are

binary encoded.

38. (Previously presented) The computer implemented method of claim 25, wherein n is at least two.

39. (Previously presented) The computer implemented method of claim 25, wherein each query sequence of the plurality of query sequences and the target sequence fragments are divided into an even number of query sequence segments and target sequence fragment segments, and further wherein the query sequence segments and the target sequence fragment segments are distributed in equal numbers between the first query group and the second query group and the first target group and a second target group.

40. (Previously presented) The computer implemented method of claim 25, wherein the n+1 query sequence segments are formed from a splitting of each query sequence of the plurality of query sequences.

41. (Previously presented) The computer implemented method of claim 25, wherein the n+1 query sequence segments are formed from a coding or scrambling of each query sequence of the plurality of query sequences.

42. (Previously presented) The computer implemented method of claim 25, further comprising: using a hash function to split each query sequence of the plurality of query sequences and the target sequence fragments into prefixes and suffixes.

43. (Previously presented) The computer implemented method claim 25, further comprising: constructing or computing at least one hash function table.

44. (Currently amended) An apparatus for searching ~~genetic data or~~ biochemical sequence information, comprising:

a memory device, wherein the memory device stores information for searching for a plurality of query sequences in a set of target sequence fragments, allowing for mismatches at up to n sequence positions;

a processing device,

wherein the processing device constructs a plurality of query distribution schemes by dividing a query sequence into a plurality of first query groups and a plurality of second query groups for storage in the memory device,

~~wherein the processing device constructs, for each query sequence of the plurality of query sequences, a first query group and a second query group, by dividing each query sequence of the plurality of query sequences into $n+1$ query sequence segments and distributing the query sequence segments between the first query group and the second query group in one or more ways such that at least n query sequence segments are contained in each second query group, and~~

~~further wherein the processing device constructs a first target group for each target sequence fragment in the set of target sequence fragments,~~

wherein the processing device constructs a plurality of corresponding target distribution schemes by dividing the target fragments into a plurality of first target groups and a plurality of second target groups for storage in the memory device; and

~~further wherein one or more of the first target groups have query sequence segment distributions corresponding to the first query groups, and~~

~~further wherein the processing device, for each query and target distribution scheme query sequence of the plurality of query sequences, compares one or more~~

of the first query groups with ~~one or more~~ of the first target groups and identifies potential matching query sequences and matching target sequence fragments for storage in the memory device;

wherein for each of the potential matches of first query groups and first target query groups, the processing device further compares the second query group and the second target group to identify a match; and

an output device for outputting information regarding the matching query sequences and ~~the matching~~ target sequence fragments.

45. (Previously presented) The apparatus of claim 44, wherein the processing device constructs, for each query sequence of the plurality of query sequences and for each target sequence fragment, respectively, a plurality of first query groups and second query groups having distinct distributions of segments between the first query groups and the second query groups.

46. (Previously presented) The apparatus of claim 44, wherein the processing device constructs, at least for each target sequence fragment identified by the processing device as a matching target sequence fragment of the matching target sequence fragments, one or more second target groups, wherein one or more of the second target groups has a segment distribution corresponding to the second query groups, and further wherein the processing device, for each identified matching query sequence and for each matching target sequence fragment, compares a corresponding second query group and a second target group to identify a match allowing for mismatches in up to n sequence positions.

47. (Previously presented) The apparatus of claim 44, wherein the processing device, for each distinct distribution of a query sequence segment, constructs a first query table indexed by possible values of the first query groups, wherein entries in the first query table provide access to each second query group by using as an index a value of a corresponding first query group.

48. (Previously presented) The apparatus of claim 47, wherein the processing device, for each distinct distribution of a query sequence segment, constructs a second query table, wherein the second query table provides access to each second query group, and further wherein the entries in the first query table provide reference to appropriate entries in the second query table.

49. (Previously presented) The apparatus of claim 44, wherein the processing device forms the $n+1$ query sequence segments by splitting each query sequence of the plurality of query sequences.

50. (Previously presented) The apparatus of claim 44, wherein the processing device forms the $n+1$ query sequence segments by coding or scrambling each query sequence of the plurality of query sequences.

51. (Previously presented) The apparatus of claim 44, wherein the processing device uses a hash function to split the plurality of query sequences and the target sequence fragments into prefixes and suffixes.

52. (Previously presented) The apparatus of claim 44, wherein the processing device constructs or computes at least one hash function table.

53. (Previously presented) The apparatus of claim 44, wherein the apparatus is a personal computer or a desk top computer.

54. (Currently amended) A computer program for installation and operation in a device for searching ~~genetic data or~~ biochemical sequence information, said device for searching comprising a memory device, a processing device, a storage device and an output device, said computer program comprising a computer program code for searching for a plurality of query

sequences in a set of target sequence fragments, allowing for mismatches at up to n sequence positions, said computer program code comprising:

a first program module or portion for constructing in said memory device, query distribution schemes for each query sequence of the plurality of query sequences, a first query group and a second query group by dividing each query sequence of the plurality of query sequences into n+1 query sequence segments and distributing the query sequence segments between the first query group and the second query group in one or more ways such that at least n query sequence segments are contained in each second query group;

a second program module or portion for constructing in said memory device, target distribution schemes for each target sequence fragment of the set of target sequence fragments, one or more first target groups having target sequence fragment segment distributions corresponding to the first query groups; and

a third program module or portion for comparing, for each query sequence of the plurality of query sequences, one or more first query groups with corresponding one or more first target groups to identify a potential match;

a fourth program module or portion for comparing, for each identified potential match, one or more second query groups with corresponding one or more second target groups and to output a result identifying matching query sequences and matching target sequence fragments.

55. (Previously presented) The computer program of claim 54, wherein the computer program is stored on a removable computer-readable storage medium.

56. (Previously presented) The computer program of claim 54, wherein the computer program is loadable on said device for searching, over a network connection.

57. (Previously presented) The computer implemented method of claim 25, wherein each query sequence of the plurality of query sequences and the target sequence fragments comprise polypeptide sequence data.